Spike Detection Using the Continuous Wavelet Transform

Zoran Nenadic*, *Member, IEEE* and Joel W. Burdick, *Member, IEEE*

Abstract—**This paper combines wavelet transforms with basic detection theory to develop a new unsupervised method for robustly detecting and localizing spikes in noisy neural recordings. The method does not require the construction of templates, or the supervised setting of thresholds. We present extensive Monte Carlo simulations, based on actual extracellular recordings, to show that this technique surpasses other commonly used methods in a wide variety of recording conditions. We further demonstrate that falsely detected spikes corresponding to our method resemble actual spikes more than the false positives of other techniques such as amplitude thresholding. Moreover, the simplicity of the method allows for nearly real-time execution.**

Index Terms—**Arrival time estimation, continuous wavelet transform, unsupervised spike detection.**

I. INTRODUCTION

ELECTRICAL recordings of action potentials have become an indispensable tool in neuroscience. The shapes and amplitudes of these action potentials, or spikes, are highly stereotyped [1]. Since such regularity obviously carries very little information, it has long been argued that the information capacity of a spike train is largely dominated by the temporal variability of individual spikes within the train [2]. It is further debated whether the nervous system cares about the precise timings of individual spikes (time coding) or the total number of action potentials fired in a certain time window (rate coding) [3]. Either way, for experimental investigations it is ultimately important to accurately and robustly detect and localize the occurrence of individual spikes within the extracellular recording signal. Essentially, all studies of the information content of experimental recordings start from this processed data. Errors in detecting the number and location of spikes will necessarily propagate through all subsequent analyses.

Extracellularly recorded spike trains are inevitably corrupted by noise. The noise sources are quite varied: the recording hardware, the ambient recording environment, the superimposed activity of multiple neurons, and the spatially averaged activity of distant cells also known as the local field potential. Perhaps most importantly, the activity of distant neurons may appear as noise which is highly correlated with the useful signal. Another difficulty is that unlike their intracellular counterparts, extracellular potentials have the shapes and amplitudes that are highly variable. These variations are influenced by many factors, most notably the cell geometry, the distribution and density of individual ionic channels, and the position of a recording electrode with respect to electrically active membranes [4]. All of these issues make the problem of spike detection challenging.

Because of its practical importance to experimental neuroscience, the detection of spikes in noisy extracellular observations is a classical problem. There exist dozens of algorithms that accomplish the task of spike detection and they can be classified as manual and automated, supervised and unsupervised. As discussed below, we are particularly interested in spike detection methods that are automated and unsupervised.

The most common manual spike detection tool is the window discriminator [5].[1] Extracellular signals that exceed a simple amplitude threshold and pass through a subsequent pair of user-specified time-voltage boxes are identified as spikes. Although practically successful, this method requires human supervision and its manual nature makes it especially time consuming when using multiple electrodes. Also, the statistical properties of this method are not well understood. Another widely used technique for spike detection is amplitude thresholding, where the threshold value can be set automatically (e.g., as a multiple of the estimated noise standard deviation), or manually. While this detection method is simple, its performance deteriorates rapidly under low signal-to-noise ratio (SNR) conditions. Other detection methods include power detection [6], [7], matched filtering [7], principal components [8], Haar transformation [9], and the discrete wavelet packet transform [10]. In the power detection method, also known as energy detection [11], the instantaneous power of the signal, calculated using a sliding window approach, is compared against a threshold derived from the mean and the standard deviation of the noise power. While this method usually offers some improvement over the amplitude threshold method, it performs almost as poorly in a low SNR environment. A matched filter (generalized matched filter) approach is known to maximize SNR when a signal is embedded in a white (colored) noise. However, because it is based on template matching, this method cannot be applied in an unsupervised fashion—supervision is required to construct the templates. The method of principal components requires the construction of the spike autocorrelation matrix, where multiple spikes from different cells are collected before the actual detection procedure. Generally, methods based on template matching or prior knowledge of "typical" spike shape are supervised detection algorithms, and not the subject of this paper. Yang and

*Z. Nenadic is with the Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: zoran@caltech.edu).

J. W. Burdick is with the Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125 USA.

[1]The primary purpose of window discriminators is to classify spikes originating from different cells. In order to be classified, spikes need to be detected first.

Shamma [9] proposed a spike detection method using the discrete Haar transformation, which is essentially a wavelet idea. However, our wavelet method does not require their white noise assumption and unnecessary inverse transformation from wavelet domain to time domain. Oweiss [10] proposed a multiresolution version of the generalized likelihood ratio test (GLRT) for spike detection, a method somewhat similar in spirit to the one developed here. The major differences between the two methods will be pointed out later. It suffices to say that the multiresolution version of GLRT developed in [10] cannot be implemented in an unsupervised manner.

Our interest in unsupervised methods arises from our current efforts to develop control algorithms that autonomously reposition extracellular recording electrodes so as to optimize and then maintain high signal recording quality in the face of inherent electrode and cell migrations in neural tissue [12]. To assess recording quality in the feedback loop, action potentials must be isolated on an ongoing basis. Because autonomously moving electrodes will experience displacements of hundreds of microns, the shape, phase, and amplitude of the detected spikes will vary significantly over the electrodes' movement ranges. Also, the moving electrode is likely to record from different cells with different action potentials along its movement track. Consequently, template-based methods are inappropriate for our use. Moreover, within the range of the electrode's movement, the recorded signal may experience widely varying SNR conditions, ranging from low-noise—high-signal amplitude near the axon hillock, to high-noise—low-signal amplitude when no neurons are present nearby. Such a wide "dynamic range" requires an unsupervised spike detection algorithm with robust performance over a wide range of parameters, which, to our knowledge, none of the existing unsupervised algorithms is capable of. The algorithm presented in this paper provides this capability in an unsupervised fashion. Moreover, our performance evaluation (Section III) shows that this approach performs better than traditional methods in many situations and, therefore, can be profitably applied in situations where traditional methods are currently used. The paper is organized as follows. In Section II we use the theory of continuous wavelet transform combined with basic detection and estimation theory to develop a new neural transient detection algorithm. Section III evaluates performance of the algorithm with respect to other commonly used spike detection methods. The results were cross-validated using data synthesized from actual recording experiments. Concluding remarks are given in Section IV. Some of the mathematical derivations are given in the Appendix. The MATLAB™ code of our method and a supporting tutorial are available at: http://robotics.caltech.edu/~zoran/Research/detection.html.

II. SPIKE DETECTION VIA WAVELETS

Our methodology consists of a combination of several techniques stemming from multiresolution wavelet decomposition, statistics, detection theory and estimation theory. For convenience, we state the five major steps of the algorithm up-front. Each step will be explained in subsequent sections.

1) Perform multiscale decomposition of the signal using an appropriate wavelet basis.
2) Separate the signal and noise at each scale.
3) Based on results from steps 2) and 3), perform Bayesian hypothesis testing at different scales to assess the presence of spikes.
4) Combine the decisions at different scales.
5) Estimate the arrival times of individual spikes.

The chances of detecting a signal embedded in noise are improved when one can take advantage of prior information about the signal and the noise. The prior information could be acquired through biophysical considerations or experimental trials. To keep the algorithm general and unsupervised, the prior information must be as vague as possible. This will be discussed in Sections II-A and II-B. Two assumptions about the noise are used in the derivations throughout this paper, the background noise is: a) stationary; b) Gaussian. Although these assumptions are not crucial for implementation, they insure the mathematical tractability of the derivations. When they are violated, the performance of the algorithm may be different from the results reported in Section III. The stationarity of neural noise cannot be assumed in general. For time scales considered in this article we assume that the statistics of the noise does not change appreciably. If the stationarity ever becomes a concern, the data can be broken into shorter segments that can be analyzed separately. Also, wavelet bases of compact support are well suited for representation of nonstationary signals. In Section III-A we shall give a justification of the Gaussian assumption.

The problem of detecting transients in a collection of noisy observations has been studied for decades. The presence of a useful signal in a background noise is normally cast as a hypothesis testing problem, where no signal is present under the null hypothesis [11]. If the signal to be detected is not perfectly known, which is the case for unsupervised problems, usually no uniformly most powerful test exists [13]. In these cases, the performance of a detector depends on the signal representation. A signal's representation can be model based or expansion based. When no appropriate model for the signal can be found, one normally projects the signal onto a canonical set of basis functions, which gives rise to a set of expansion coefficients. Depending on the signal representation, the detection problem can be formulated in a variety of domains: the time domain, frequency domain, time-frequency domain, etc.

In this paper we use a time-scale expansion based signal representation. As we will show, there exist wavelet basis functions that are well suited to spike detection. Moreover, there exist fast algorithms for wavelet filtering [14], thereby allowing real-time implementation. A brief introduction to the continuous wavelet transform is presented next. For more details on this subject, interested readers can consult [15] or other texts. For a look at other approaches to signal detection using wavelet bases, see [16]–[18]. These techniques are based on dyadic wavelets (see Section II-B), and are mainly concerned with the detection of a single transient within the observation sequence, without the estimation of parameters such as transient arrival times.

A wavelet ψ is a function of finite energy, i.e., $\psi \in L^2(\mathbf{R})$, and zero average

$$\int_{\mathbf{R}} \psi(t) \, dt = 0.$$

It is normalized $\|\psi\|_2 = 1$ and centered in the neighborhood of the origin. From this function, also called the *mother wavelet*,

one can obtain a family of time-scale waveforms by translation and scaling

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \quad a,b \in \mathbf{R} \tag{1}$$

where $a > 0$ represents the scale and b is the translation. The functions $\psi_{a,b}$ are called wavelets and they share the properties of the mother function [note that $\psi = \psi_{1,0}$, so that the mother function is a member of the family defined by (1)].

The *wavelet transform* of an arbitrary function $x \in L^2(\mathbf{R})$ is a projection of that function onto the wavelet basis

$$Tx(a,b) = \int_{\mathbf{R}} x(t)\psi_{a,b}(t)\,dt. \tag{2}$$

For a fixed scale a_0 and translation b_0, the wavelet transform of the function x, denoted by $X(a_0, b_0)$, represents its resemblance index to the wavelet ψ_{a_0,b_0}. If the index is "large" the resemblance is strong and vice versa. This index is termed the wavelet coefficient. For wavelets of compact support [19], the domain of integration in (2) is confined to the support. Consequently, the wavelet coefficient only depends on the part of the signal within the support. If the scale of the wavelet function is relatively small, the wavelet of compact support provides a local analysis tool suitable for capturing transient signal phenomena and for coping with nonstationaries.

A. Wavelet Functions for Spike Detection

It is important to choose a wavelet that is suitable for the signal of interest. Our choice is motivated by the shape of the extracellular potentials to be detected in the background noise. This shape can be explained with the help of biophysics. Despite the differences in extracellular and intracellular potentials pointed out in Section I, they are fundamentally related. The extracellular potential depends on the transmembrane current, which consists largely of a resistive component and a capacitive component. Since the capacitive component is proportional to the time derivative of the transmembrane potential, it is approximately biphasic.[2] The capacitive component may dominate the membrane current during an action potential, so the time course of the extracellular spike is typically biphasic [4]. Because a wavelet coefficient represents the measure of similarity between the signal and the basis function, it is reasonable to look for a wavelet that is spike-like. Accordingly, the neural signal would be represented by a few coefficients. In approximation theory this is known as a *sparse representation* and the basis functions corresponding to these coefficients can be interpreted as intrinsic signal structures [15]. In the presence of noise, the sparse representation becomes an important condition for successful separation of signal and noise (see Section II-C).

Fig. 1 shows several different wavelet functions. The simplest wavelet is the *Haar* function. While it is compact, its discontinuity implies that it is not well localized in the frequency domain. The other wavelet functions shown are *db2*[3] and two wav-

[2]If the transmembrane action potential undergoes after-hyperpolarization, the capacitive current can be tri-phasic, with the third phase having significantly smaller amplitude than the rest of the waveform.

[3]This is a family of wavelets introduced by Daubechies, hence, the name. The number, 2 in this case, represents the order of wavelet and is equal to the number of vanishing moments of ψ.
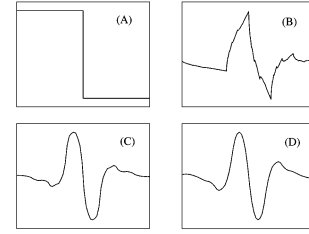

Fig. 1. Wavelets of four different wavelet families: (A) *Haar*, (B) *db2*, (C) *bior1.3*, and (D) *bior1.5*.
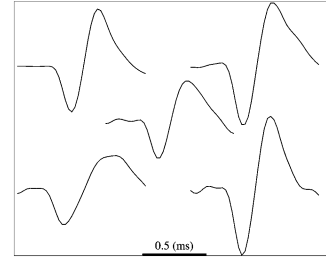

Fig. 2. Five spike templates identified from single-electrode recordings in the cortex of a Rhesus monkey. The spikes were detected, aligned and clustered (see Section III-A). The waveforms corresponding to the same class were then averaged for smoothing purposes and the five templates were obtained.

elets from the family of biorthogonal wavelets: *bior1.3* and *bior1.5*. Biorthogonal wavelets are constructed from splines and more about these wavelets can be found in [20]. We note that unlike *db2*, the biorthogonal wavelets are symmetric and that *db2* wavelets do not have explicit analytic expression. Fig. 2 shows five spike templates recorded from the posterior parietal cortex of a Rhesus monkey (*Macaca mulatta*) using a single electrode (see Section III-A for recording details). Note that the two biorthogonal wavelets appear to match the intrinsic structure of the templates better than *db2*, as their biphasic shape is reminiscent of action potentials. It is, therefore, expected that the biorthogonal wavelets provide a sparser representation of neural signal than *db2*, which together with its near symmetric variate was used in [10].

B. Choice of Scale

The continuous wavelet transform defined by (2) operates on a continuous set of scales and translations. Hence, the basis functions $\psi_{a,b}$ are not orthogonal and the representation of the signal x by its wavelet coefficients is redundant. One can choose dyadic scales and translations from a discrete set $\{a = 2^j; b = k2^j; j, k \in \mathbf{Z}\}$, so that the corresponding wavelets $\psi_{a,b}$ form an orthonormal basis of $L^2(\mathbf{R})$. Here, *we will restrict the set of scales and translations in a different manner*. Practically, all extracellularly recorded signals are sampled in discrete time. Thus, we choose the set of basis function translations to be finite, where this set is determined by the sampling rate of the signal $f_s(\mathrm{kHz})$ and its duration $T(\mathrm{s})$, i.e., $b \in \mathcal{B}$, where

$$\mathcal{B} = \{0, 1, \ldots, k, \ldots, N-1\}$$

and $N = Tf_s + 1$ is the number of samples of the discrete signal (time series). Therefore, in the continuous wavelet transform[4] the set of translations coincides with the discrete time vector.

Similarly, biophysical considerations of the duration of extracellular potentials can be used to restrict the relevant scales of the wavelet basis functions. Despite their variability in shape and amplitude, the vast majority of extracellular spikes are highly localized in time, with a characteristic duration. For example, action potentials in primate cortex typically last for 0.4–0.5 ms for signals recorded near the axon, and 0.7–1.0 ms for recordings near the soma-dendritic complex [4]. Similar results have been reported in [21], although depending on the species and the brain area, action potentials can last up to 3.0 ms. Based on this biophysical knowledge of the duration of the transient to be detected, we can limit the set of scales for the analyzing wavelet functions. This practically serves to filter out a significant amount of noise and also appreciably reduce the real-time computational requirements. In summary, we use a limited set of scales

$$\mathcal{A} = \{a_0, a_1, \ldots, a_j, \ldots, a_J\}$$

where a_0 and a_J are determined from the signal sampling rate and the minimum and maximum spike durations, denoted by $W_{\min}$ and $W_{\max}$, respectively, and these two parameters are chosen by the user. In contrast to the conventional dyadic scaling used in most wavelet applications, we choose the intermediate scales $\{a_1, a_2, \ldots, a_{J-1}\}$ uniformly sampled between the two extrema a_0 and a_J with an arbitrary step chosen by the user. In the case of dyadic scaling, the scales of wavelet decomposition range from very fine to very coarse, where these limits are determined by the signal duration and sampling rate. The coefficients at very fine scales contain nothing but signal discontinuities and high frequency noise and are not relevant for detection of neural transients. Likewise, at coarser scales the information about relatively short transients is contaminated through the excessively large support of the basis functions. This is another major difference between our approach and the one proposed by [10] that is based on dyadic scaling. Hereafter, the wavelet coefficient of discrete signal x at scale $a_j \in \mathcal{A}$ and translation $k \in \mathcal{B}$, is denoted by $X(j,k)$, where $X(j,k) = \langle x, \psi_{j,k} \rangle$, and $\langle , \rangle$ stands for the inner product in $\mathbf{R}^N$.

C. The Statistics of Wavelet Coefficients

By applying the continuous wavelet transform, with possibly a restricted set of scales and translations, we obtain a multiscale representation of the signal in terms of its wavelet coefficients. If the discrete observations x contain useful signal s and noise w, i.e.,

$$x[n] = s[n] + w[n] \quad n \in \mathcal{B} \tag{3}$$

then the statistical properties of the wavelet coefficients will depend on those of the noise. For example, if w is a white Gaussian noise (WGN) with mean μ and variance σ^2, it follows from (3) and the properties of the wavelet transform that the mean and the variance of the coefficient $X(j,k)$ are

$$E\{X(j,k)\} = S(j,k)$$

and

$$\text{Var}\{X(j,k)\} = \sigma^2$$

where $S(j,k) = \langle s, \psi_{j,k} \rangle$ is the wavelet coefficient of the true signal s. If the representation of the signal is sparse, only a few

of the expected coefficient values will be different from zero. In other words, the coefficients $X(j,k)$ corresponding to noise are zero-mean random fluctuations, and the coefficients $X(j,k)$ corresponding to "signal plus noise" are random variables with means different from zero.

For purposes of unsupervised signal detection, we must separate these coefficients by estimating the noise level σ in each coefficient from the sampled data. To obtain these estimates, we borrow ideas from Donoho and Johnstone [22] who studied the problem of nonlinear estimation of signals from noisy data under a sparse representation. Their remarkably good wavelet denoising method is based on accepting only those wavelet coefficients that exceed a threshold, followed by the inverse wavelet transform. In our case, the threshold will become a part of the hypothesis testing procedure at the level of coefficients (see Section II-D), without the need for an inverse transformation. They proposed a hard thresholding rule

$$\rho_T(X) = \begin{cases} X, & \text{if } |X| > T \\ 0, & \text{if } |X| \leq T \end{cases} \tag{4}$$

which acts on the wavelet coefficients X at different scales, and was shown to perform close, with respect to a certain metric, to an ideal estimator. For a near-optimal performance it is sufficient to choose the threshold $T = \sigma \sqrt{2 \log_e N}$, where, as before, N is the number of samples of the analyzed time series. If the background noise is not white, the threshold becomes basis function dependent [15]. In this case, each wavelet coefficient generally has a threshold of its own.

From the translation invariance of the continuous wavelet transform and the underlying noise, it follows that the coefficients at scale a_j have the same threshold, i.e., $T_j = \sigma_j \sqrt{2 \log_e N}$, where σ_j^2 is the variance of the noise coefficients $W(j,k)$ at scale a_j. Since these coefficients are not known, this variance must be estimated from the observation coefficients $X(j,k)$. Due to a possible presence of useful signal s in the observation x, such an estimate will be biased. That is, the values of $X(j,k)$ significantly different from $W(j,k)$ will contain useful signal. However, such coefficients will be outliers because few coefficients contain signal due to our sparse representation. The sample variance of the set of coefficients at scale a_j

$$\mathcal{X}_j \triangleq \{X(j,k) : k \in \mathcal{B}\}$$

will be affected by those outliers, in particular it will over-estimate σ_j^2. From robust statistics [23], we know that the median of a random variable is less sensitive to outliers than its variance. For a Gaussian random variable (see Appendix I) it can be shown that the median of its absolute deviation effectively estimates the standard deviation

$$\hat{\sigma}_j = M\{|X(j,0) - \bar{\mathcal{X}}_j|, \ldots, |X(j,N-1) - \bar{\mathcal{X}}_j|\}/0.6745 \tag{5}$$

where $\bar{\mathcal{X}}_j$ is the sample mean of $\mathcal{X}_j$ and $M\{\cdot\}$ denotes the sample median. Equation (5) was derived under the assumption that $X(j,k)$ are independent Gaussian random variables, which may not hold in general. The Gaussian nature of the coefficients follows from the Gaussian noise assumption and the linearity of the wavelet transform. Furthermore, for a rich class of $1/f$ processes, it has been shown that the wavelet decomposition at a fixed scale nearly whitens (decorrelates) coefficients at that

scale [24]. Under the Gaussian assumption, this implies that the sequence $\mathcal{X}_j$ is independent. Fig. 3 shows a colored noise sequence and its continuous wavelet representation for several different scales,[5] together with the corresponding auto-covariance sequences (ACVS). It is apparent that the ACVS of the wavelet coefficients at different scales resemble the Dirac function, indicating that the coefficients are uncorrelated. Also note that the white noise approximation is less valid at larger scales, due to a significant amount of overlap in the basis functions. This can be circumvented by subsampling the sequence of wavelet coefficients at scale a_j, say $\mathcal{X}_j = \{X(j,k) : k = 0, a_j, 2a_j, \ldots\}$ which makes the resulting sequence nearly white and the estimate given by (5) approximately valid. The estimate $\hat{\sigma}_j$ is then used in the detection algorithm that will be presented next.

D. The Detection Algorithm at a Single Scale

We now formulate the detection problem in the time-scale domain. The problem of detecting spikes in a noisy signal can be seen as a binary hypothesis testing problem, where under the null hypothesis $\mathcal{H}_0$ the signal is not present, and under the alternative $\mathcal{H}_1$ both signal and noise are present. More formally

$$\mathcal{H}_0 : x[n] = w[n] \quad n \in \mathcal{B}$$
$$\mathcal{H}_1 : x[n] = s[n] + w[n] \qquad n \in \mathcal{B}$$

where $x[n]$ represents a noisy observation (evidence) at a discrete time n, s is the transient (spike) to be detected and w is the background noise. Because of the transient nature of s, the hypothesis $\mathcal{H}_1$, if true, will be so only for an interval of time, or equivalently for a subset of the discrete time $\mathcal{B}$. Moreover, multiple transients could be present, and these represent the main differences between the problems of the classical signal detection and the detection of action potentials.

We formulate the first step of our detection problem as a sequential binary hypothesis test at each scale $a_j \in \mathcal{A}$. Section II-E discusses how to combine the coefficient level decisions. Appendix II derives the following hypothesis testing rule for each wavelet coefficient $X(j,k)$

$$|X(j,k)| \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \frac{\hat{\mu}_j}{2} + \frac{\hat{\sigma}_j^2}{\hat{\mu}_j} \log_e \gamma_j \triangleq \Theta_j \quad \forall k \in \mathcal{B} \qquad (6)$$

where $\hat{\sigma}_j$ is determined from (5), $\hat{\mu}_j$ is the sample mean of the absolute value of the wavelet coefficients at scale a_j under the hypothesis $\mathcal{H}_1$, γ_j is a parameter that depends upon the acceptable costs of false alarms and omissions, denoted by λ_{FA} and λ_{OM}, and the prior probabilities of the two hypotheses, denoted by $P(\mathcal{H}_0)$ and $P(\mathcal{H}_1)$ (see below and Appendix II for more details). The parameter Θ_j can be viewed as an acceptance threshold for the hypothesis $\mathcal{H}_1$ at scale a_j.

Note that $\hat{\mu}_j$ cannot be estimated from $\mathcal{X}_j$, as many of the coefficients from $\mathcal{X}_j$ will contain noise only. However, we can get a rough estimate as to which coefficients contain noise only, by applying the thresholding defined by (4), where the value of the threshold is determined by $T_j = \hat{\sigma}_j \sqrt{2 \log_e N}$. Such a procedure splits $\mathcal{X}_j$ into two disjoint subsets: a noise subset $\mathcal{X}_j^w$ and a signal subset $\mathcal{X}_j^s$, where

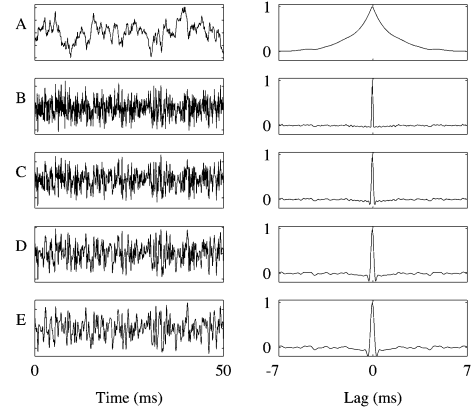$$\mathcal{X}_j^w \triangleq \{X(j,k) \in \mathcal{X}_j : |X(j,k)| \leq T_j\}$$

Fig. 3. (A) (left) Colored process obtained as a steady-state response of a first-order linear system (time constant 1.3 ms, sampling rate 20 kHz) driven by a zero-mean WGN of unit variance [25]. (right) The normalized ACVS of the process showing correlation up to several milliseconds, consistent with the time constant of the system. (B)–(E) (left) Continuous wavelet transform of the process at scales 2, 4, 6, and 8, respectively. (right) The normalized ACVS of the corresponding multiscale representation of the process indicating significant correlation only in the vicinity of zero lag.

and

$$\mathcal{X}_j^s \triangleq \mathcal{X}_j \setminus \mathcal{X}_j^w.$$

Thus, $\hat{\mu}_j$ is found as a sample mean of the absolute value of the elements of $\mathcal{X}_j^s$, and $\hat{\mu}_j \triangleq 0$ if $\mathcal{X}_j^s = \{\emptyset\}$.

In order to evaluate γ_j we must specify the costs λ_{OM} and λ_{FA}. The ratio $P(\mathcal{H}_0)/P(\mathcal{H}_1)$ of two prior[6] probabilities is determined as $\|\mathcal{X}_j^w\|/\|\mathcal{X}_j^s\|$, where $\|\cdot\|$ stands for the size of a set. Also note that γ_j depends only on the ratio of λ_{FA} and λ_{OM} (see Appendix II), and we can constrain the two costs by $\lambda_{\text{FA}} + \lambda_{\text{OM}} = 1$. In that case γ_j can be conveniently reparametrized, i.e.,

$$\log_e \gamma_j = \log_e \left[\frac{\lambda_{\text{FA}}}{1 - \lambda_{\text{FA}}} \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} \right] \triangleq LL_M + \log_e \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} \qquad (7)$$

where L is in effect a new parametrization of $\lambda_{\text{FA}}/\lambda_{\text{OM}}$ and $L_M = 36.7368$ is a conveniently chosen scaling factor. The choice of L_M is not unique and the value above corresponds to the natural logarithm of the maximum ratio of λ_{FA} and λ_{OM} that does not cause arithmetic overflow under a double-precision floating point representation. In this reparametrization, a relatively narrow range of L conveniently covers a relatively wide range of false alarm-to-omission cost ratios. For example, $L \in [-0.188, 0.188]$ corresponds to $\lambda_{\text{FA}}/\lambda_{\text{OM}} \in [0.001, 1000]$, with $L = 0$ equivalent to $\lambda_{\text{FA}} = \lambda_{\text{OM}}$. Thus, for all practical applications it is sufficient to chose $L \in [-0.2, 0.2]$.

Once L is selected (which is equivalent to the choice of $\lambda_{\text{FA}}/\lambda_{\text{OM}}$), the acceptance threshold Θ_j is determined from (6) and a decision is made at each scale a_j. Section II-E describes how the decisions at different scales are combined. Before we proceed, let us closely examine an unlikely, but possible, scenario that could happen in the outcome of the thresholding outlined above.

---

[5]The nomenclature for the scales of the continuous wavelet transform used throughout this article is consistent with that of MATLAB™ Wavelet Toolbox.

[6]Since these probabilities are estimated from the data, they are not priors in the strict Bayesian probability sense.
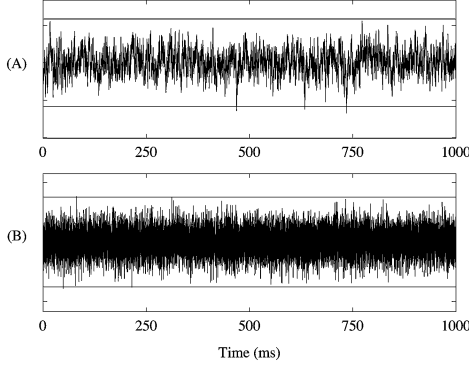
Fig. 4. (A) Colored noise sequence generated in the same way as in Fig. 3. (B) White noise sequence. The horizontal lines mark the 3.75 standard deviation bounds.

TABLE I
THE NUMBER OF DETECTED TRANSIENTS FOR DIFFERENT CHOICE OF
DETECTION PARAMETER USING TWO DIFFERENT DETECTION METHODS:
WAVELET (CONSERVATIVE AND LIBERAL) AND AMPLITUDE THRESHOLDING

Method	Parameter	Colored	White
Wavelet (conservative)	any L	0	0
Amplitude thresholding	$T_s = 3.00$	11	65
	$T_s = 3.25$	8	27
	$T_s = 3.50$	5	8
	$T_s = 3.75$	3	3
	$T_s = 4.00$	3	0
Wavelet (liberal)	$L \in [0.0, 0.2]$	0	0

The estimate of the acceptance threshold Θ_j for the hypothesis $\mathcal{H}_1$ is data dependent, and largely depends on how the set $\mathcal{X}_j$ splits into two subsets $\mathcal{X}_j^w$ and $\mathcal{X}_j^s$. What happens in the degenerate case when $\mathcal{X}_j^s = \{\emptyset\}$? Such a case occurs when no wavelet coefficients exceed the threshold T_j, and is conceivable if the hypothesis $\mathcal{H}_0$ is true $\forall n \in \mathcal{B}$, i.e., observations contain noise only. Fig. 4 shows two such sequences containing only Gaussian noise (colored and white). If $\mathcal{X}_j^s = \{\emptyset\}$, it follows from (6) and (7) that $\Theta_j = \infty$ regardless of L, and we expect our detection algorithm to accept $\mathcal{H}_0$ sequentially $\forall k \in \mathcal{B}$, thereby not detecting any transients. Other detection methods, such as amplitude thresholding, are not capable of recognizing if observations contain any useful signal or noise only.

To illustrate this point, Table I shows the number of events detected from the two random sequences shown by Fig. 4 using two different detection methods: our wavelet method, whose description will be completed in Section II-E, and amplitude thresholding. The detection based on our method is parameter independent ($\Theta_j = \infty$) and produces no detected transients. The detection based on amplitude thresholding is performed for several different values of the threshold T_s determined by the number of standard deviations of noise. Table I clearly shows that amplitude thresholding suffers from a number of false positives, even for conservative estimates of the threshold based on 3.75 or 4 standard deviations of the noise. This feature of the wavelet detection algorithm may be useful if one is to avoid false positives, and this mode of operation is termed "conservative".

Although the conservative mode is desirable regarding the probability of false alarm, it may be inadequate for some applications. Namely, if the observations contain a very weak transient signal (low SNR), the coefficients of useful signal and noise are rather similar, and the algorithm may refuse to detect any transients. The reason for this is that none of the coefficients will survive the hard thresholding given by (4). Consequently the set $\mathcal{X}_j^s$ will be empty and the acceptance threshold of $\mathcal{H}_1$ will never be attained.

Alternatively, we can slightly modify the algorithm so that it works in a "liberal" acceptance mode. In this case, we perturb the parameters comprising the acceptance threshold Θ_j. Since $P(\mathcal{H}_1) = 0$ under $\mathcal{X}_j^s = \{\emptyset\}$, we replace this probability by the smallest nonzero probability of $\mathcal{H}_1$, which is $P(\mathcal{H}_1) = 1/N$. This is equivalent to the assumption that there is a single wavelet coefficient at scale a_j that exceeds the threshold T_j. Under this assumption we know that the sample mean of $\mathcal{X}_j^s$ is at least T_j, therefore, we choose $\hat{\mu}_j = T_j$. These will render the acceptance threshold Θ_j finite and, hence, the mode of operation is termed "liberal." Therefore, if $\mathcal{X}_j^s = \{\emptyset\}$, the liberal mode parameters can be viewed as a local perturbation of the conservative mode parameters. If $\mathcal{X}_j^s \neq \{\emptyset\}$, the two modes are equivalent. The bottom part of Table I shows that the performance of the algorithm is usually not compromised using liberal mode for a wide range of values of parameter L, and we will be using this mode of operation through the rest of this paper.

E. The Overall Detection Algorithm

This section describes our overall spike detection and estimation algorithm. The methodology consists of two major steps.
  i) Combine the decisions at individual scales.
  ii) Estimate the spike arrival times.

1) Combining Decisions at Individual Scales: Because they are highly localized in time, the samples corresponding to neural transients occupy contiguous subsets of the discrete time vector $\mathcal{B}$. This property of transients is often referred to as a *temporal contiguity*. Temporal contiguity translates into the contiguity of the coefficients in the wavelet domain [26], i.e., the wavelet coefficients corresponding to the same transient tend to be neighbors in both time and scale. Since we use the continuous wavelet transform with the basis functions of compact support roughly matched to the scale of neural transients, the temporal contiguity in the wavelet domain is inherently preserved. The scale contiguity follows from a broad frequency spread of a time-limited signal, namely if a scale is thought of as an approximation of the frequency, a time-limited transient will be spread across many scales. The presence of noise, however, may obscure the picture at the scales that are not relevant. The scale contiguity can also be viewed in the present context as a cross-correlation (redundancy) of the wavelet coefficients (decisions) at different scales. Before we turn to the problem of redundancy, let us introduce a few notations.

Let $\mathcal{B}_j^{\mathcal{H}_1}$ be a subset of the translation set $\mathcal{B}$ that corresponds to the acceptance of $\mathcal{H}_1$ at scale a_j, i.e., $\mathcal{B}_j^{\mathcal{H}_1} \triangleq \{k \in \mathcal{B} : |X(j,k)| > \Theta_j\}$. If nonempty, the sets $\mathcal{B}_j^{\mathcal{H}_1}$ comprise a number of contiguous regions, where a *contiguous region at scale a_j* is defined as the subset of the translation set $\mathcal{B}$ over which the hypothesis $\mathcal{H}_1$ is accepted in succession at the scale a_j. These concepts are illustrated by Fig. 5, which shows a segment of a simulated spike train, the wavelet coefficients that support the acceptance of $\mathcal{H}_1$, and the corresponding translation indices defining
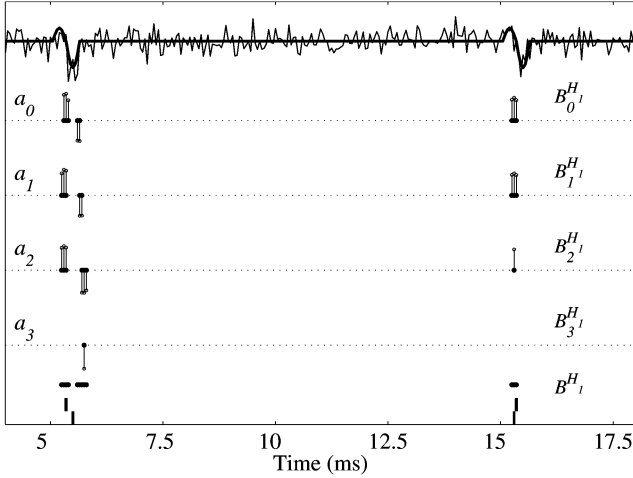
Fig. 5. A segment of a spike train with two transients. The spikes (black solid line) with known occurrence times are modeled as two semi-cycles of a sine wave with different amplitudes. For simplicity, the signal is corrupted by a zero-mean WGN and the resulting signal is shown in gray. The subsequent plots show the wavelet coefficients at scales $\{a_0, a_1, a_2, a_3\}$ that support the acceptance of $\mathcal{H}_1$ and the corresponding contiguous regions that define sets $\mathcal{B}_j^{\mathcal{H}_1}$, as well as the set $\mathcal{B}^{\mathcal{H}_1}$, obtained by (8). The tick marks represent the true arrival times of spikes (black) and their estimates (gray).

$\mathcal{B}_j^{\mathcal{H}_1}$ together with its contiguous constituents. The scales are chosen assuming $W_{\min} = 0.5$ ms and $W_{\max} = 0.8$ ms, so that the support of the basis functions at scales $a_0, a_1, a_2,$ and a_3 are 0.5, 0.6, 0.7, and 0.8 ms, respectively. The duration of the noise-free transients is taken to be 0.6 ms, so we expect the transients to be best detected at scale a_1. Note that at each scale $a_0, a_1,$ and a_2 there are three contiguous regions, and that the information provided at these scales is highly redundant. Also note that the first transient is hardly detected at scale a_3 and that the second one is omitted, which is not surprising given that the most relevant scale for the transient considered is a_1, and that the degradation in performance increases as the scales deviate from a_1.

The redundancy is lost by defining the combined set of acceptance of $\mathcal{H}_1$, denoted by $\mathcal{B}^{\mathcal{H}_1}$, as

$$\mathcal{B}^{\mathcal{H}_1} = \bigcup_{a_j \in \mathcal{A}} \mathcal{B}_j^{\mathcal{H}_1}. \tag{8}$$

By applying (8), the contiguities across different scales are combined. We note that such a procedure only makes sense in the case of the continuous wavelet transform because the basis functions at different scales are defined over the same set of translations $\mathcal{B}$. Therefore, if $\mathcal{B}^{\mathcal{H}_1} \neq \{\emptyset\}$ it comprises a number of contiguous regions, where a *contiguous region* is defined as the subset of the translation set $\mathcal{B}$ over which the hypothesis $\mathcal{H}_1$ is accepted in succession at any of the analyzing scales. The combined set of acceptance of $\mathcal{H}_1$ containing three contiguous regions is shown at the bottom of Fig. 5.

2) Estimation of Spike Arrival Times: We now turn to the issue of estimating the spike arrival times from the wavelet coefficients supporting the acceptance of $\mathcal{H}_1$. In a noise-free environment, the wavelet basis function that provides the maximum correlation with the transient to be detected, corresponds to a wavelet coefficient of maximum magnitude. The time associated with the translation index of the basis function with maximal coefficient can be taken as a good approximation to the

occurrence time of the underlying transient. Because we choose the set of translations $\mathcal{B}$ with time resolution down to the sampling period, this approximation is essentially as good as the sampling period. Tracking of modulus maxima of the wavelet coefficients across scales has been proposed for the detection of signal singularities [27]. In a noisy environment, there is naturally a jitter associated with the location of this maximal coefficient. This jitter can be reduced by averaging the locations of the maxima across different scales. This is the idea behind our approach.

To estimate the spike arrival times, we start by organizing the acceptance set of $\mathcal{H}_1$ into its contiguous constituents, i.e.,

$$\mathcal{B}^{\mathcal{H}_1} = \bigcup_{i=1}^{N_c} \mathcal{C}_i^{\mathcal{H}_1}$$

where $\mathcal{C}_i^{\mathcal{H}_1}$ are the contiguous regions of $\mathcal{B}^{\mathcal{H}_1}$ and N_c is the number of contiguous regions. Let T_i^j be the estimated location of the ith transient at scale a_j, i.e., it is the translation index of the maximum magnitude coefficient in the ith contiguous region at scale a_j

$$T_i^j \triangleq \arg \max_{k \in \mathcal{C}_i^{\mathcal{H}_1}} \{|X(j,k)| : |X(j,k)| > \Theta_j\}$$

$$\forall i = 1, 2, \ldots, N_c \quad \forall a_j \in \mathcal{A}.$$

Note that T_i^j may not exist for all scale-contiguous region pairs. For example, T_1^3 and T_3^3 are not defined in Fig. 5, as no coefficient at scale a_3 supports the acceptance of $\mathcal{H}_1$ over the sets $\mathcal{C}_1^{\mathcal{H}_1}$ and $\mathcal{C}_3^{\mathcal{H}_1}$. The arrival time *candidate* of the ith transient is then found by averaging T_i^j over those scales that support the acceptance of $\mathcal{H}_1$ on the ith contiguous region

$$T_i = \frac{1}{\|\mathcal{A}_i^{\mathcal{H}_1}\|} \sum_{a_j \in \mathcal{A}_i^{\mathcal{H}_1}} T_i^j \quad \forall i = 1, 2, \ldots, Nc$$

where $\mathcal{A}_i^{\mathcal{H}_1} \triangleq \{a_j \in \mathcal{A} : |X(j,k)| > \Theta_j, k \in \mathcal{C}_i^{\mathcal{H}_1}\}$, and, as before, $\|\cdot\|$ is the size of a set. In the example given by Fig. 5, $\mathcal{A}_1^{\mathcal{H}_1} = \{a_0, a_1, a_2\}, \mathcal{A}_2^{\mathcal{H}_1} = \{a_0, a_1, a_2, a_3\}$, and $\mathcal{A}_3^{\mathcal{H}_1} = \{a_0, a_1, a_2\}$. Before the candidate T_i is declared as the arrival time of an individual transient, an additional processing step may be necessary.

First, note that as a consequence of the multiphasic shape of wavelet functions (see Fig. 1) and the fine resolution of the translation set $\mathcal{B}$, the coefficients corresponding to a single transient will smoothly vary over both positive and negative values in a subset of the translation set $\mathcal{B}$. The smallest coefficients will vanish after the threshold operation (6) is applied, resulting in a distribution of the nonzero wavelet coefficients over two (or more) contiguous regions. An example of this phenomenon can be found in the first transient of Fig. 5, which has two corresponding contiguous regions. However, the disjoint contiguous regions that arise from a single transient will be very close in time, and the spurious transients associated with this phenomenon can be reliably eliminated by merging the candidates that are sufficiently close in time, followed by a re-estimation of the arrival time based on the merged intervals. This type of postprocessing is performed sequentially from the onset of the signal for two candidates at a time. The intermediate results are updated and the process is repeated until no further merging is possible. The arrival time of the first transient from Fig. 5, is estimated in this fashion, where the candidates T_1 and T_2 are combined to produce a single arrival time.
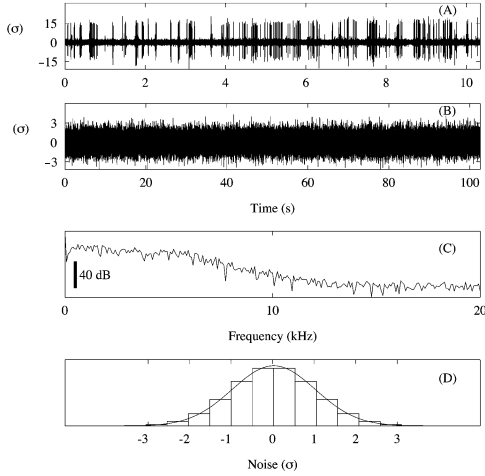
Fig. 6. (A) An extracellularly recorded signal with high SNR. (B) A signal containing no distinguishable spikes-neural noise. The amplitude in two panels is shown in σ units, where σ is the standard deviation of the neural noise. (C) Power spectrum of the neural noise. (D) Histogram of the neural noise together with a Gaussian pdf with the same mean and variance as those of the noise samples.

The only remaining unknown is how close two candidates need to be in order to be combined. The answer to this question depends on the largest scale a_J used in the analysis, as the distance between two arrival times belonging to the same transient tend to increase with the scale (see Fig. 5). Recall that the coarsest scale a_J is determined from the maximal duration of the transient (see Section II-B), therefore, two transients will be merged if their distance is less than $W_{\max}$, where $W_{\max}$ is the maximum duration of the transient to be detected.

This method, therefore, successfully eliminates the spurious candidates that belong to a single spike. The obvious drawback, however, is that we are unable to resolve two genuine transients that are closer than $W_{\max}$. For a well-isolated single unit recording, this is not a serious constraint, as two spikes cannot be generated arbitrarily close in time due to the refractory period [28]. However, spikes that originate from different units can partially overlap. Thus, the postprocessing of the spike candidates should be viewed as a tradeoff between the elimination of the spurious transients and inability to resolve partially overlapped spikes.

Once the spike arrival times are estimated, the spikes are extracted by copying a fixed number of samples before and after the estimated arrival. This results in a vector representation of individual events, convenient for further analysis. We proceed by testing the proposed algorithm in a realistically modeled detection task. Throughout the rest of the article, we will refer to this algorithm as the wavelet detection method (WDM).

III. SIMULATION RESULTS

Analytical assessment of the performance of the detector (6) requires a full knowledge of the statistics of the wavelet coefficients under both hypotheses. In the absence of this knowledge, one can resort to Monte Carlo simulations in order to evaluate the detector performance. To ensure consistent results, the number of Monte Carlo trials needs to be sufficiently high and individual trials have to be independent. It is all but impossible to properly test the detection performance on real extracellular

signals, as we do not have independent information about the number of spike transients and their exact arrival times in each trial. The only reliable way to obtain this information is to perform simultaneous intracellular recording [6], [29], which is virtually impossible in behaving animals. Hence, to rigorously test the performance of this method, we use simulations that synthesize spike trains from actual recorded data.

A. Modeling Neural Data

One way to model neural data is to extract spikes from actual recordings, arrange them randomly in time with an arbitrary firing rate and corrupt them by a suitably modeled "neural noise," where the noise level is determined by the value of SNR chosen for study. Fig. 6 shows a high SNR extracellular data set collected[7] from the posterior parietal cortex of a Rhesus monkey. A good separation of signal and noise in this data enables reliable spike detection. Transients are detected from the data using the WDM with $L = 1.0$. This value of the parameter imposes an extremely high penalty for false alarms $\lambda_{\text{FA}}/\lambda_{\text{OM}} = \exp(37.7368)$, and ensures a low probability of Type-I error (see Appendix II). The detected spikes $\mathcal{S} = \{S_1, S_2, \ldots, S_{N_s}\}$ were then aligned using a maximum correlation method. Briefly, the spikes are normalized $\|S_n\|_2 = 1$, and the first detected spike is fixed. The second spike is then shifted locally in time to maximize its correlation with the first spike. The third spike is now shifted back and forth until the sum of its correlations with the first two spikes is maximal and so on. A similar alignment procedure has been proposed by [8]. This procedure effectively re-estimates the arrival time each detected spike. Some $N_s = 285$ spikes were detected and aligned in this manner. The spikes were scaled back to their original size and their peak-to-peak amplitudes were calculated. From the scatter plot of the peak-to-peak amplitudes, five different clusters were visually identified. The spikes were clustered using the k-means method [30] and the cluster averages were calculated, resulting in five templates which are shown in Fig. 2. We use multiple spike template shapes in our simulations to ensure that the algorithm is not overly sensitive to a particular waveform shape. The same template construction procedure was repeated using amplitude thresholding detection with the threshold set so that 285 spikes are detected. The resulting templates did not exhibit significant deviations from the ones shown in Fig. 2. Therefore, the use of WDM in the construction of simulation templates does not impose any bias in the further analysis.

These templates were used to generate spike trains with known properties as follows. Spike arrival times were generated by a homogeneous Poisson process with a certain firing rate (FR) and a refractory period of 2 ms enforced. The simulation was terminated once the number of arrival times reached a prespecified number N_a. The five templates were then randomly drawn with equal relative frequencies and were successively centered at the arrival times, barring boundary

---

[7]A single platinum-iridium microelectrode (Frederic Haer Company, Bowdoinham, ME) with the nominal impedance of 2 MΩ at 1 kHz was used for the recording. The data was acquired through a recording system (Plexon Inc, Dallas, TX) with a preamplifier and a band-pass filter (band 154 Hz–13 kHz). The signals were amplified and digitized (12 bit A/D converter, digitization rate 40 kHz) by a data acquisition card [PCI-MIO 16E-4 with LabView (National Instruments, Austin, TX)].

conditions. Fig. 6 also shows a monkey extracellular data record (recorded as above) that did not contain any visible spikes and was, therefore, used as a template for neural noise.[8] The subsequent panels show the periodogram (power spectrum) and the histogram of this neural noise signal. Note the characteristic $1/f$ trend in the spectrum and bell-shaped noise histogram indicating nearly Gaussian trend in the data. Gaussian approximation of background neural noise has been discussed in [31], [32]. Each spike within the spike train was normalized so that $\|S_i\|_\infty = 1$, and a randomly selected subsegment of the neural noise was scaled according to the SNR desired for each test. That is, the standard deviation of the neural noise was defined according to

$$\sigma \triangleq \frac{\|S_i\|_\infty}{\mathrm{SNR}} = \frac{1}{\mathrm{SNR}} \qquad (9)$$

and added to the generated spike train. This procedure was repeated for different FR and SNR values. For each choice of FR and SNR, many trials were performed. Also note that random arrangements of the spike templates effectively remove any systematic bias.

B. A Note on SNR

In communication theory, SNR at time t is defined as the amplitude of the signal at time t divided by the standard deviation of the noise [33]. Such a time dependent definition is not particularly useful in neurophysiology, where SNR can be viewed as a single number that characterizes the noisiness of a spike train. Colloquially, SNR can be defined as the ratio of the maximum amplitude of the spike and the "amplitude" of the background noise, hence, the definition (9). If a spike train contains multiunit activity, different spikes are likely to have different SNRs, thus, we need to apply averaging in order to obtain a single SNR. The "average" SNR may be biased depending on the relative frequency of spikes of different amplitudes, and we avoid this problem by normalizing the spike templates. Another commonly used definition of SNR involves the root-mean-square values of the spike and noise [7], and can be recovered from (9) under the Euclidean norm and a scale factor. This definition of SNR gives lower values than the one given by (9), under identical noise variance. Finally, the definition of SNR may involve the power of signal and noise as opposed to their amplitudes. This discrepancy can be reconciled by expressing the SNR in decibels [$20 \log_{10}(\mathrm{SNR})$ for amplitude and $10 \log_{10}(\mathrm{SNR})$ for power]. Because of this variability in the SNR definition, it is often difficult to objectively assess the level of noise in the data. To alleviate this difficulty and give some feeling for the noise level, Fig. 7, provides a snapshot of the simulated data at two different SNRs.

C. Performance Evaluation

From the discussion of Section II-C it follows that the number of action potentials affects the estimate of the noise coefficient standard deviation σ_j, therefore, the performance of

Fig. 7. (A) Spike train at $\mathrm{SNR} = 3.5$ (left) and the blow-up of the shaded region (right). The tick marks indicate the locations of the spikes. Note how certain noise samples have amplitudes comparable to the amplitudes of spikes. (B) Equivalent plots for $\mathrm{SNR} = 4.0$.

the WDM was tested at low, medium and high firing rates of 10, 30, and 100 Hz. Because the arrival times follow the statistics of a Poisson process, the length of individual trials is variable. Setting the total number of spikes per trial to $N_a = \mathrm{FR}$, renders the average duration of trials to be approximately 1 s. Relatively high noise levels, $\mathrm{SNR} = \{3.5, 3.6, \ldots, 4\}$, were applied in the analysis.

The performance of the WDM was compared against other methods including the power detection method (PDM), the single amplitude thresholding method (SATM) (see Section I), and the double amplitude thresholding method (DATM). Since the spike polarity might change during the course of an experiment, the SATM may be inadequate (e.g., a positive threshold is bound to miss the majority of negative going spikes). For truly unsupervised applications, we introduce the DATM, where either positive or negative threshold crossings indicate the presence of spikes. The performance of these methods was assessed using the receiver operating characteristic (ROC). The ROC curve compares the probability of (correct) detection ($\mathrm{P_D}$) versus the probability of false alarm ($\mathrm{P_{FA}}$). The ROC curve for each combination of FR and SNR was obtained by averaging the performance over 300 trials. Appendix III details our averaging methodology.

A recent detection method using a nonlinear energy operator (NEO) [34] reports a successful detection under nearly 0 dB, where, the authors defined SNR as a peak-to-peak amplitude of the smallest spike template divided by the noise standard deviation. However, without specifying the amplitudes and relative frequencies of other templates, followed by averaging, as discussed in Section III-B, SNR can be manipulated to any desired value. In other words, such a definition of SNR does not provide an objective measure of the noisiness of the data, as larger amplitude spike templates will have larger SNRs. Furthermore, the authors generated neural noise using autoregressive moving average (ARMA) model, although it has been shown that ARMA models are inadequate for capturing long-term correlation structure typically associated with $1/f$ processes [24], [35]. We tested the performance of the NEO detector under a properly defined SNR and realistic neural noise model and found its performance at best comparable, and often falling short

---

[8]The boundary between signal and noise in the context of extracellular recording is not clearly defined due to the fact that a large component of noise represents the activity of distant neurons. A data record such as the one shown on Fig. 6 is, therefore, treated as noise, even though it is likely to contain low amplitude spikes.
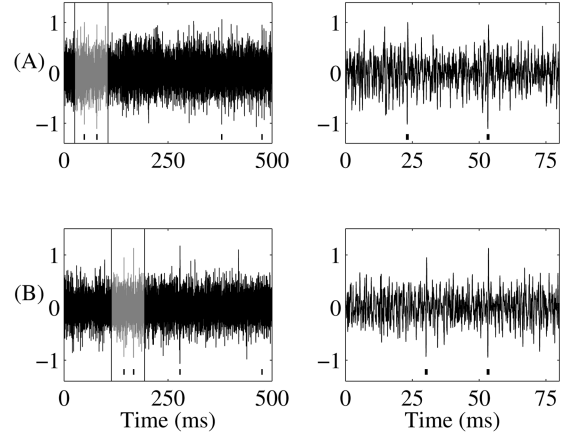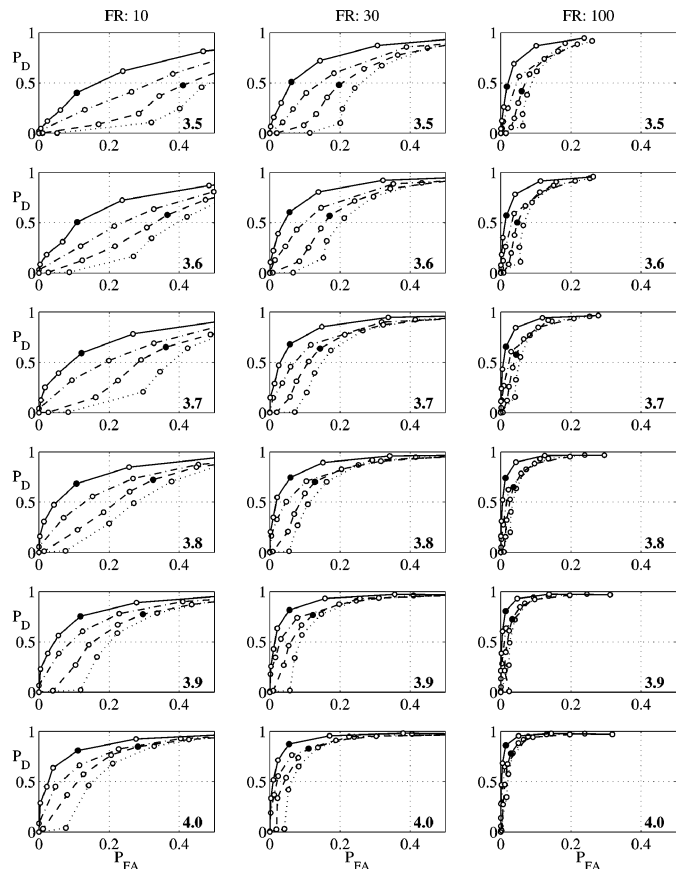
Fig. 8. ROC for different FR and SNR scenarios. The horizontal axis of each panel represents the probability of false alarm and the vertical axis represents the probability of detection. Each column represents ROC at fixed mean firing rate: low 10 Hz, medium 30 Hz, and high 100 Hz. Each row represents the performance at fixed SNR: $3.5, 3.6, \ldots, 4$ (bottom right corner of each panel). The detection methods are: WDM (—), DATM ($\ldots$), SATM (- - -), and PDM ($- \cdot - \cdot -$). Filled circles correspond to detection with $L = 0$ for WDM and $T_s = 3.6$ for SATM.

Fig. 9. Comparison of performances of WDM, SATM, DATM, and PDM with the detection parameters set at $L = 0, T_s = 3.6, T_d = 3.75$, and $T_p = 3$, respectively (see Section III-C for details). (A) Performance of WDM. The arrows show the direction of increase of FR and SNR. The dotted lines represent interpolated performance at fixed SNR and for FR between 10 Hz and 100 Hz. (B) Performance of SATM. (C) Performance of DATM. (D) Performance of PDM.

of other conventional methods. Therefore, the detection using NEO will not be a subject of further investigation.

The results of our simulation tests for different combinations of firing rate and SNR are shown in Fig. 8. Each ROC curve is obtained by a systematic variation of the threshold parameter; parameter L in the WDM and parameters T_s, T_d, and T_p, which represent a multiple of the estimated noise standard deviation, in SATM, DATM, and PDM, respectively. Clearly, the ROC corresponding to the WDM lies above the ROC of other detection methods. In other words for a fixed probability of false alarm (P_{FA}), the probability of detection (P_D) of the WDM is consistently bigger than those of the other methods. Conversely, at the same level of P_D, the WDM has consistently lower false alarm rate than the other methods. The improvement in performance is more apparent under low SNR and low FR conditions, which are commonly encountered in everyday recording practice. Also note that our results are consistent with the results of [7] which showed that the PDM outperforms SATM. Likewise, SATM slightly outperforms DATM. This is not surprising given the near symmetric shape of the templates (see Fig. 2), making the double amplitude thresholding redundant, thereby increasing the probability of false alarm.

We also note that unlike the classical ROC which depicts the performance of a binary hypothesis test and which should lie
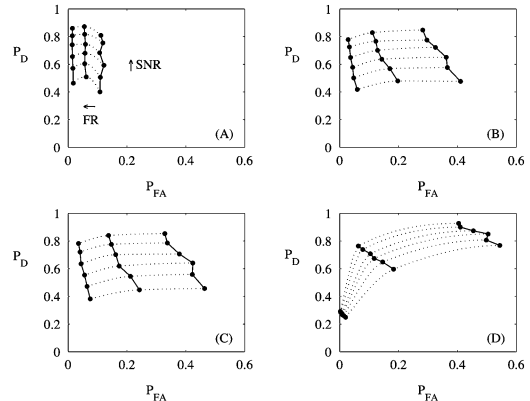
above the chance line ($P_D = P_{FA}$), the ROCs shown in Fig. 8 have a different character. First, they correspond to the result of sequential hypothesis testing. Second and more importantly, the detected signals are transient and their arrival times are estimated. Thus, the ROCs from Fig. 8 represent combined results of detection and estimation. Since the estimation problem is more sensitive to the presence of noise than detection [11], a chance detector would result in an ROC with $P_{FA} \gg P_D$. Therefore, the performance curve of a combined detector and estimator can lie below the chance line, although this tends to happen only for SATM and DATM, and only under relatively low P_D conditions.

1) Parameter Settings for Unsupervised Applications: For unsupervised detection one does not have the luxury of varying the detection threshold, hence, a single threshold that "works" across a wide range of SNRs and FRs is sought. We investigate how choosing a single parameter affects the performance of different methods, and find that the WDM offers additional advantage over the other detection methods, as shown in Fig. 9. For approximately equal levels of P_D [Fig. 9(A) versus Fig. 9(B), (C), and (D)], we see that the dispersion of P_{FA} of the WDM is much smaller than that of the other methods. Not only does the WDM provide a smaller P_{FA}, but also it has a more consistent behavior over a range of FRs. In addition, for a fixed FR, P_{FA} remains fairly constant across different SNRs, as can be seen in the almost vertical performance curves of Fig. 9(A). On the contrary, the performance curves of the other methods tend to be slanted [Fig. 9(B)–(D)], and this becomes more apparent at low FR. We conclude this analysis by noting that the value of parameter $L = 0$ [Fig. 9(A)] offers a reasonable compromise between P_{FA} and P_D while maintaining a consistent performance over a wide range of FRs and SNRs. Therefore, for unsupervised spike detection, this might be a good choice of the parameter.

2) Timing Jitter: As discussed above, the WDM can be viewed as a combination of detection and parameter estimation, where the estimated parameters are the spike arrival times. The presence of noise, however, causes jitter in these estimates (see Fig. 5). Two important performance parameters of any estimator are its bias and consistency. Ideally, an estimator should

TABLE II
BIAS AND STANDARD DEVIATION OF ESTIMATED SPIKE ARRIVAL TIMES FOR
WDM AND SATM AVERAGED OVER TRIALS AND DIFFERENT SNR
AND FR SCENARIOS

	WDM		SATM	
	Jitter (ms)	Std (ms)	Jitter (ms)	Std (ms)
Average	-0.0396	0.0633	-0.1523	0.0497

Fig. 10.(A) Features (first two principal components) of false positives of WDM clustered in 4 distinct groups ($\bullet$, $\blacktriangle$, $+$, and $\square$). White circles indicate the features of the five templates. (B) Equivalent plot for SATM, where data tend to fall into a single cluster ($\bullet$), with one outlier ($+$). (C) The waveforms (cluster averages) corresponding to clusters from (A). (D) Equivalent plot for cluster from (B) and the outlier.

be unbiased with consistency proportional to the number of observations. Table II shows estimated jitter and its standard deviation averaged over 300 trials and over different SNR-FR scenarios.

The results correspond to WDM with $L = 0$ and SATM with $T_s = 3.6$. The negative value of jitter indicates that the estimated spike times are lagging the true spike times. On average, the WDM has a bias that is nearly 4 times smaller than that of SATM. On the other hand, the standard deviation of the jitter is smaller in the case of SATM, and this can be viewed as a bias-variance tradeoff. Despite their presence, bias and inconsistency do not pose a serious threat in the process of spike detection for several reasons. First, their values are rather small compared to the duration of spike transients and they are comparable to the sampling period of 0.025 ms of the data. Second, they can be eliminated if desired through the process of spike alignment (see Section III-A), which effectively re-estimates the occurrence times of individual events.

3) Characteristics of WDM False Positives: Finally, we performed an analysis of the falsely detected spikes under the $\mathrm{WDM}(L = 0)$ and $\mathrm{SATM}(T_s = 3.6)$ to assess the nature of the failure modes of the two approaches. Our conjecture is that the false positives of WDM will resemble true spikes more than the false positives of SATM. Fig. 10 summarizes the results for $\mathrm{SNR} = 3.5$ and $\mathrm{FR} = 10$ Hz, although similar behavior was observed over all SNR-FR combinations. The false positives over 300 trials for the two methods were extracted. Falsely identified spikes were then represented in a two-dimensional feature space by means of principal components. Although projected to the same subspace, the false positives corresponding to WDM [Fig. 10(A)] and SATM [Fig. 10(B)] are shown on separate plots for clarity. The false positives corresponding to WDM tend to cluster in 4 distinct groups, where those of SATM fall into a single broad cluster. The features of the five spike templates used for data modeling are also shown in these plots. Fig. 10(C) and (D) shows the cluster average waveforms, so that each spike within the cluster can be treated as a noisy realization of the corresponding waveform. The majority of the falsely detected spikes by the WDM are biphasic ($\bullet$ and $\blacktriangle$ waveforms) with a small number of tri-phasic spikes ($+$ and $\square$ waveforms), where most of the false positives detected by the SATM are mono-phasic and do not have the realistic spike shape. Furthermore, the spike templates fall into one of the clusters of WDM, and this is not surprising given the shapes of the templates (Fig. 2) and the shape of $\bullet$ waveform [Fig. 10(C)]. This happened consistently for all SNR-FR combinations. Therefore, the false positives of the WDM are well structured and could represent the activity of distant neurons, where the false positives of the SATM mostly represent random voltage fluctuations. Hence, one can generally make a good use of false positives of WDM where false positives of SATM
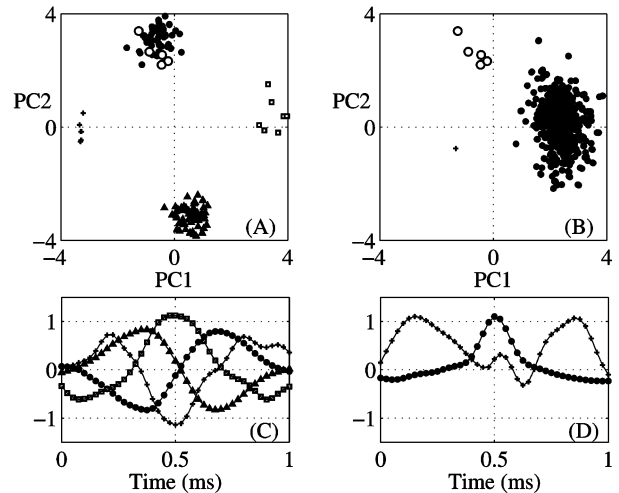
have to be discarded as noise. For real applications, this is done through the process of spike clustering [31], [36].

IV. CONCLUSION

Despite the large number of existing algorithms for detection of extracellular potentials in noisy observations, robust, fully automated detection algorithms have been scarce. We presented a novel detection scheme and compared its performance to many commonly used spike detection methods. The detection is cast in the standard hypothesis testing framework and since the signals to be detected are unknown, the detector performance is representation dependent. Spike transients have historically been detected by a simple amplitude thresholding, where the threshold level is chosen with respect to the (estimated) standard deviation of the noise. We have shown how this can yield erroneous results, especially if no signal is present. Spike waveforms are not just samples whose average amplitude exceeds some baseline level. They also have a characteristic shape and duration. Using wavelets we are taking advantage of this additional information that is ignored by amplitude or power thresholding methods.

The algorithm is completed by combining detection, which arises from sequential hypothesis testing, and parameter estimation, where the occurrence times of individual spike transients are the parameters of interest. Additional postprocessing of estimated arrival times may be necessary. This may result in inability to resolve the transients that are within a millisecond of each other. However, the same problem is inherent to any other detection method, and there have been some attempts to provide solutions [7], [36], although in the context of spike classification rather than spike detection.

The proposed algorithm has been extensively tested over a wide range of conditions via Monte Carlo simulations of spike trains that were synthesized from actual data. We found that the WDM consistently outperforms other common detection methods, and the differences have been quantified. The basic

findings are: a) the WDM provides a significant improvement under extremely low SNRs and low FRs, a situation commonly found in actual experiments; s) for a single choice of detection parameter, the WDM offers more consistent performance under different SNR-FR scenarios; 3) therefore, it is possible to come up with a single parameter that performs well for a wide range of SNRs and FRs, which is very useful for unsupervised on-line applications; 4) the jitter in estimated spike arrival times and its variance are comparable for WDM and commonly used amplitude thresholding; 5) falsely detected spikes of WDM are likely to be caused by the activity of distant neurons and can be utilized as neural data, as opposed to false positives of amplitude thresholding methods, which typically represent random fluctuations (noise).

We conclude this paper by noting that the WDM is also suitable for off-line analysis, where by trading off the cost of omission and false alarm, the user can modify the sensitivity of the method, similar to changing the threshold level in amplitude thresholding method.

APPENDIX I
MEDIAN ABSOLUTE DEVIATION

Let X be a Gaussian random variable with mean μ and variance σ^2, denoted by $X \sim \mathcal{N}(\mu, \sigma^2)$. Define random variables $Y \triangleq X - \mu$ and $Z \triangleq |Y|$. Clearly $Y \sim \mathcal{N}(0, \sigma^2)$ and the probability density function (pdf) of Z can be written as

$$f_Z(z) = \frac{dP_Z(Z \le z)}{dz} \tag{10}$$

where $P_Z(Z \le z)$ is the cumulative distribution function of Z. It follows from the definition of Z that

$$P_Z(Z \le z) = \begin{cases} P(-z \le Y \le z) & z \ge 0 \\ 0 & z < 0 \end{cases}. \tag{11}$$

Combining (10) and (11) yields

$$f_Z(z) = \begin{cases} \dfrac{d}{dz} \displaystyle\int_{-z}^{z} f_Y(y)\,dy = 2f_Y(z), & z \ge 0 \\ 0, & z < 0 \end{cases} \tag{12}$$

where the result follows after differentiating the integral from above with respect to the upper and lower bounds. The median of Z, denoted by $M\{Z\}$, can be defined as the middle point of the pdf $f_Z(z)$, i.e.,

$$\int_{-\infty}^{M\{Z\}} f_Z(z)\,dz \triangleq \frac{1}{2}. \tag{13}$$

By substituting (12) into (13) and by noting that $f_Y(z) = (2\pi\sigma^2)^{-1/2}\exp(-z^2/2\sigma^2)$ one obtains

$$M\{Z\} = \sigma\sqrt{2}\,\mathrm{erf}^{-1}\left(\frac{1}{2}\right)$$

where erf stands for the error function defined by $\mathrm{erf}(x) = 2(\pi)^{-1/2}\int_0^x \exp(-t^2)\,dt$, and erf^{-1} represents its inverse. After numerical evaluation it follows that $M\{Z\} \approx 0.6745\sigma$. The significance of this result is that one can use the median of a sample for estimation of its standard deviation. Since the median is much less susceptible to the presence of outliers, it provides a more robust estimate of the standard deviation. Let $\mathcal{X} = \{x_i\}_{i=1}^N$ be a sequence of N independent identically distributed Gaussian random variables with variance σ_j^2. From the analysis above, it follows that

$$\hat{\sigma} = M\{|x_1 - \bar{\mathcal{X}}|, |x_2 - \bar{\mathcal{X}}|, \dots, |x_N - \bar{\mathcal{X}}|\}/0.6745$$

where $M\{\mathcal{S}\}$ is the sample median of the sequence $\mathcal{S}$ and $\bar{\mathcal{X}} \triangleq 1/N \sum_i x_i$ is the sample mean of $\mathcal{X}$.

APPENDIX II
SEQUENTIAL BINARY HYPOTHESIS TESTING

Recall that we formulate our detection problem as a binary hypothesis testing problem, where under the null hypothesis $\mathcal{H}_0$ the signal is not present, while under hypothesis $\mathcal{H}_1$ the signal is present

$$\mathcal{H}_0 : x[n] = w[n] \quad n = 0, 1, \dots, N-1$$
$$\mathcal{H}_1 : x[n] = s[n] + w[n] \quad n = 0, 1, \dots, N-1$$

where $x[n]$ represents a noisy observation at time n, s is the signal to be detected and w is noise. By the linearity of wavelet transform at scale $a_j \in \mathcal{A}$, these two hypotheses have the following form:

$$\mathcal{H}_0 : X(j, k) = W(j, k) \quad k \in \mathcal{B}$$
$$\mathcal{H}_1 : X(j, k) = S(j, k) + W(j, k) \quad k \in \mathcal{B}.$$

As in any hypothesis testing problem, the goal is to determine whether the evidence supports the rejection of $\mathcal{H}_0$. This decision should be made optimally with respect to a suitably chosen objective function. Let $\mathcal{R}(\mathcal{H}_0 \,|\, X)$ and $\mathcal{R}(\mathcal{H}_1 \,|\, X)$ be conditional risks associated with accepting and rejecting the hypothesis $\mathcal{H}_0$ given the evidence X, respectively. These risks can be expressed as

$$\mathcal{R}(\mathcal{H}_0 \,|\, X) = \lambda_{00} P(\mathcal{H}_0 \,|\, X) + \lambda_{01} P(\mathcal{H}_1 \,|\, X)$$
$$\mathcal{R}(\mathcal{H}_1 \,|\, X) = \lambda_{10} P(\mathcal{H}_0 \,|\, X) + \lambda_{11} P(\mathcal{H}_1 \,|\, X)$$

where $\lambda_{ij} \ge 0$ is the cost of accepting $\mathcal{H}_i$ given that the true state of nature is $\mathcal{H}_j$ and $P(\mathcal{H}_i \,|\, X)$ is the probability of accepting $\mathcal{H}_i$ given the evidence. It is customary not to penalize for correct decisions, therefore, $\lambda_{00} = \lambda_{11} = 0$. Furthermore, λ_{10} represents the cost of rejecting $\mathcal{H}_0$ when it is true (Type-I error or false alarm) and λ_{01} is the cost of accepting $\mathcal{H}_0$ when it is not true (Type-II error or omission error). Hence, we adopt the following notation: $\mathcal{R}(\mathcal{H}_0 \,|\, X) = \lambda_{\mathrm{OM}} P(\mathcal{H}_1 \,|\, X)$ and $\mathcal{R}(\mathcal{H}_1 \,|\, X) = \lambda_{\mathrm{FA}} P(\mathcal{H}_0 \,|\, X)$. The overall Bayes risk is then defined as

$$\mathcal{R} = \int_\Omega [\mathcal{R}(\mathcal{H}_0 \,|\, X) + \mathcal{R}(\mathcal{H}_1 \,|\, X)] p(X) dX \tag{14}$$

where $p(X)$ is the pdf of X. The optimal decision rule that minimizes the cost (14), is to accept the hypothesis with a smaller conditional risk [37], i.e.,

$$\mathcal{R}(\mathcal{H}_0 \,|\, X) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \mathcal{R}(\mathcal{H}_1 \,|\, X). \tag{15}$$

That is, (15) is read as: "accept $\mathcal{H}_1$ if $\mathcal{R}(\mathcal{H}_0 \,|\, X) > \mathcal{R}(\mathcal{H}_1 \,|\, X)$, and vice versa." After invoking the Bayes rule $P(\mathcal{H}_i \,|\, X) = p(X \,|\, \mathcal{H}_i) P(\mathcal{H}_i)/p(X)$, the decision rule (15) becomes

$$\frac{p(X \,|\, \mathcal{H}_1)}{p(X \,|\, \mathcal{H}_0)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \frac{\lambda_{\mathrm{FA}}}{\lambda_{\mathrm{OM}}} \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} \triangleq \gamma, \tag{16}$$

where $p(X \mid \mathcal{H}_i)$ is the likelihood of X given $\mathcal{H}_i$ and $P(\mathcal{H}_i)$ is the prior probability of the hypothesis $\mathcal{H}_i$. Note that γ represents the acceptance threshold for $\mathcal{H}_1$ and in the typical case that $\lambda_{\mathrm{FA}} > \lambda_{\mathrm{OM}}$ and $P(\mathcal{H}_0) \gg P(\mathcal{H}_1)$, we have that $\gamma \gg 1$. Under the assumption that the noise is Gaussian we have $p(X \mid \mathcal{H}_0) \sim \mathcal{N}(0, \sigma^2)$ and and $p(X \mid \mathcal{H}_1) \sim \mathcal{N}(\pm\mu, \sigma^2)$, where $\mu > 0$, and $\pm$ correspond to positive and negative coefficients, respectively. By the Gaussian assumption one can rewrite (16) as

$$|X| \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \frac{\mu}{2} + \frac{\sigma^2}{\mu} \log_e \gamma \tag{17}$$

where μ is the mean value of $|X|$ under the hypothesis $\mathcal{H}_1$ and σ is the standard deviation of X. The parameters μ and σ, as well as the prior probabilities $P(\mathcal{H}_0)$ and $P(\mathcal{H}_1)$ are not known, and have to be estimated from the data. In this case, the test (16) becomes essentially a GLRT and $|X|$ represents a sufficient statistic. The only difference between GLRT and Bayes decision is in the way the acceptance threshold γ is chosen. The condition (17) is checked sequentially ($\forall k \in \mathcal{B}$) at each scale $a_j \in \mathcal{A}$.

APPENDIX III
AVERAGING OVER TRIALS

For the ith simulation trial, $\mathrm{P_D}$ and $\mathrm{P_{FA}}$ are estimated as

$$\hat{P}_{\mathrm{D}}^{(i)} = \frac{N_{cd}^{(i)}}{N_a^{(i)}}$$

and

$$\hat{P}_{\mathrm{FA}}^{(i)} = \frac{N_{fa}^{(i)}}{N_d^{(i)}}$$

where $N_{cd}^{(i)}, N_a^{(i)}, N_{fa}^{(i)}$, and $N_d^{(i)}$ are the number of correctly detected spikes, generated spikes, falsely detected spikes, and detected spikes in the ith trial, respectively. A correct detection is declared if the detected arrival time is within 0.5 ms of the true arrival time. The overall estimate is obtained as an average over many trials, i.e., $\hat{P}_{\mathrm{D}} = \langle \hat{P}_{\mathrm{D}}^{(i)} \rangle$ and $\hat{P}_{\mathrm{FA}} = \langle \hat{P}_{\mathrm{FA}}^{(i)} \rangle$. Since $N_a^{(i)}$ is constant across trials, the first average simply becomes

$$\hat{P}_{\mathrm{D}} = \frac{1}{N_T} \sum_{i=1}^{N_T} \hat{P}_{\mathrm{D}}^{(i)}$$

where N_T is the total number of trials ($N_T = 300$ in our case). However, the number of detected spikes $N_d^{(i)}$ is not constant over trials, therefore, the trials cannot be given uniform weights in the estimate of $\hat{P}_{\mathrm{FA}}$. Since $\hat{P}_{\mathrm{FA}}^{(i)}$ is conditioned upon $N_d^{(i)}$, we write based on the law of total probability

$$\hat{P}_{\mathrm{FA}} = \sum_{k=0}^{\infty} \hat{P}_{\mathrm{FA} \mid \mathrm{k}} P(N_d = k)$$

where $\hat{P}_{\mathrm{FA} \mid k}$ is the estimate of $\mathrm{P_{FA}}$ given that the number of detected spikes is k, and $P(N_d = k)$ is estimated as the number of trials with k detected spikes divided by N_T. We also note that

$$\hat{P}_{\mathrm{FA} \mid k} = \begin{cases} \frac{\bar{N}_{\mathrm{fa} \mid k}}{k}, & k \in \mathbf{N} \\ 0, & k = 0 \end{cases}$$

where $\bar{N}_{\mathrm{fa} \mid k}$ is the average number of false alarms given that the number of detected spikes is k.

REFERENCES

[1] F. Delcomyn, *Foundations of Neurobiology*. New York: Freeman, 1998.
[2] S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek, "Entropy and information in neural spike trains," *Phy. Rev. Let.*, vol. 80, pp. 197–200, 1998.
[3] C. Koch, *Biophysics of Computation*. New York: Oxford Univ. Press, 1999.
[4] D. R. Humphrey, *Electrophysiological Techniques*. Atlanta, GA: Soc. Neurosci., 1979.
[5] J. P. Welsh and C. Schwarz, "Multielectrode recording from the cerebellum," in *Methods for Neural Ensemble Recordings*, M. A. L. Nicolelis, Ed. Boca Raton, FL: CRC, 1999, ch. 5, pp. 79–100.
[6] K. D. Harris, D. A. Henze, J. Csicsvari, H. Hirase, and G. Buzsaki, "Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements," *J. Neurophysiol.*, vol. 84, pp. 401–414, 2000.
[7] I. N. Bankman, K. O. Johnson, and W. Schneider, "Optimal detection, classification, and superposition resolution in neural waveform recordings," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 8, pp. 836–841, Aug. 1993.
[8] M. Abeles and M. H. Goldstein, "Multispike train analysis," *Proc. IEEE*, vol. 65, pp. 762–773, 1977.
[9] X. Yang and S. A. Shamma, "A totally automated system for the detection and classification of neural spikes," *IEEE Trans. Biomed. Eng.*, vol. 35, no. 10, pp. 806–816, Oct. 1988.
[10] K. G. Oweiss, "Multiresolution Analysis of Multichannel Neural Recordings in the Context of Signal Detection, Estimation, Classification and Noise Suppression," Ph.D. thesis, Univ. Michigan, Ann Arbor, 2002.
[11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
[12] J. G. Cham, E. A. Branchaud, B. Greger Branchaud, Z. Nenadic, R. A. Andersen Nenadic, and J. W. Burdick, "A semi-chronic motorized microdrive and control algorithm for autonomously isolating and maintaining optimal extracellular action potentials," *J. Neurophysiol*, DOI: 10.1152/jn.00369.2004, to be published.
[13] H. L. Van Trees, *Detection, Estimation and Modulation Theory*. New York: Appleton & Lange, 1991.
[14] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal Machine Intell.*, vol. 11, no. 7, pp. 674–693, July 1989.
[15] ——, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic, 1999.
[16] M. Frisch and H. Masser, "The use of the wavelet transform in the detection of an unknown transient signal," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 892–897, Feb. 1992.
[17] N. Erdol, "Multiscale detection," in *Wavelets, Subband and Block Transforms in Communications and Media*, A. N. Akansu and M. M. Medley, Eds. Boston, MA: Kluwer Academic Publishers, 1999, ch. 6, pp. 183–205.
[18] T. T. Liu and C. Fraser-Smith, "Detection of transients in $1/f$ noise with the undecimated discrete wavelet transform," *IEEE Trans. Signal Processing*, vol. 48, no. 5, pp. 1458–1462, May 2000.
[19] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. Pure Appl. Math.*, vol. 41, pp. 909–996, 1988.
[20] ——, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
[21] R. Lemon, *Methods for Neuronal Recording in Conscious Animals*. New York: Wiley, 1984, ch. 2, pp. 17–38.
[22] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
[23] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
[24] P. Flandrin, "Wavelet analysis and synthesis of fractional brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 910–917, Feb. 1992.
[25] K. J. Åström, *Introduction to Stochastic Control Theory*. New York, NY: Academic Press, 1970.
[26] Z. Wang and P. K. Willett, "All-purpose plug-in power-law detectors for transient signals," *IEEE Trans. Signal Processing*, vol. 49, no. 11, pp. 2454–2466, Nov. 2001.

[27] S. Mallat and W. L. Hwang, "Singularity detection and processign with wavelets," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 617–643, Feb. 1992.

[28] E. R. Kandel, J. H. Schwartz, and T. M. Jessel, *Principles of Neural Science*, CT: Wiley, 1981.

[29] M. Wehr, J. S. Pezaris, and M. Sahani, "Simultaneous paired intracellular and tetrode recordings for evaluating the performance of spike sorting algorithms," *Neurocomputing*, vol. 26–27, pp. 1061–1068, 1999.

[30] G. A. F Seber, *Multivariate Observations*. New York: Wiley, 1984.

[31] M. Sahani, "Latent Variable Models for Neural Data Analysis," Ph.D. thesis, California Inst. Technol., Pasadena, CA, 1999.

[32] M. S. Fee, P. P Mitra, and D. Kleinfeld, "Variability of extracellular spike waveforms of cortical neurons," *J. Neurophysiol.*, vol. 76, pp. 3823–3833, 1996.

[33] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 2002.

[34] K. H. Kim and S. J. Kim, "Neural spike sorting under nearly 0-dB signal-to-noise ratio using nonlinear energy operator and artificial neural network classifier," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 10, pp. 1406–1411, Oct. 2000.

[35] G. W. Wornell, "Wavelet-based representation for the $1/f$ family of fractal processes," *Proc. IEEE*, vol. 81, no. 10, pp. 1428–1450, 1993.

[36] M. S. Lewicki, "Bayesian modeling and classification of neural signals," *Neural Computation*, vol. 6, pp. 1005–1030, 1994.

[37] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

Zoran Nenadic (M'04) received the B.S. degree in control engineering from the University of Belgrade, Belgrade, Yugoslavia, in 1995 and the M.S. and D.Sc. degrees in systems science and mathematics from Washington University, St. Louis, MO, in 1998 and 2001, respectively.

He is currently a Postdoctoral Fellow with the Division of Engineering and Applied Science at the California Institute of Technology, Pasadena. His research interests are in the area of computational neuroscience, signal processing, and applications of control theory in biological systems.

Joel W. Burdick (M'04) received the B.S degrees in mechanical engineering and chemistry from Duke University, Chapel Hill, NC, in 1981, and the M.S. and Ph.D. degrees in mechanical engineering from Stanford University, Stanford, CA, in 1982 and 1988, respectively.

Since 1988, he has been with the Division of Engineering and Applied Science at the California Institute of Technology, Pasadena, where he is currently a Professor of Mechanical Engineering and Bioengineering. His research interests include robotic locomotion, sensor-based robot motion planing, multifingered robotic grasping, neural prosthetics, and applied nonlinear control theory.